(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 923 044 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
16.06.1999 Bulletin 1999/24

(51) Int. Cl.[6]: **G06K 9/20**

(21) Application number: **98121791.2**

(22) Date of filing: **16.11.1998**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **21.11.1997 US 975466**

(71) Applicant: **Xerox Corporation**
Rochester, New York 14644 (US)

(72) Inventor:
Syeda-Mahmood, Tanveer F.
Rochester, New York 14625 (US)

(74) Representative:
Grünecker, Kinkeldey,
Stockmair & Schwanhäusser
Anwaltssozietät
Maximilianstrasse 58
80538 München (DE)

(54) Method and means of matching documents based on spatial region layout

(57) A method for matching obejcts based on spatial layout of regions based on a shape similarity model for detecting similarity between general 2D objects. The method uses the shape similarity model to determine if two obejcts are similar by logical region generation in which logical regions are automatically derived from information in the obejcts to be matched, region correspondence, in which a correspondence is established between the regions on the obejcts, pose computation in which the individual transforms relating corresponding regions are recovered, and pose verification in which the extent of spatial similarity is measured by projecting one document onto the other using the computed pose parameters. The method of the invention can be carried out in a microprocessor-based system capable of being programmed to carry out the method of the invention.

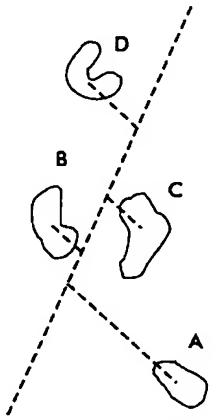

FIG. 3A

BEST AVAILABLE COPY

EP 0 923 044 A2

Description

[0001] This invention is related to object matching based on shape similarity and, more particularly, to a method and means by which spatial layout of regions can be captured for purposes of matching documents.

[0002] In today's world, increasing numbers of documents are being scanned in large quantities or are being created electronically. To maintain and manage these documents requires new methods that analyze, store and retrieve the documents. Current document management systems can support document database creation from scanned documents and indexing based on text queries. A need for allowing more visual queries has been felt, particularly in retrieving documents when text keywords are unreliably extracted (from scanned documents due to OCR errors), or retrieve too many choices for a user to select from. In such cases the intention of the user is best captured by either allowing more flexible queries making reference to a document genre or type (say, find me a "letter" from "X" regarding "sales" and "support"), or by simply pointing to an icon or example, and asking "find me a document looking similar to it in visual layout." Performing either requires an ability to automatically derive such document genre or type information from similarity in the visual layouts of documents rather than their precise text content, which may be quite different. An example illustrating this can be seen from Figures 1A and 1B which are two similar-looking documents with very different text content.

[0003] Matching based on spatial layout similarity is a difficult problem, and has not been well-addressed. The above examples also illustrate the outstanding difficulty. The two documents in Figure 1A and 1B are regarded as similar even though their logically corresponding regions (text segments) shown in Figures 2A and 2B, respectively, differ in size. Furthermore, some of the corresponding regions have moved up while others have moved down and by different amounts.

[0004] It is known to extract a symbolic graph-like description of regions and perform computationally intensive subgraph matching to determine similarity, as seen in the work of Watanabe in "Layout Recognition of Multi-Kinds of Table-Form Documents", IEEE Transactions Pattern Analysis and Machine Intelligence. Furthermore, US-A Patent No. 5,642,288 to Leung et al. entitled "Intelligent document recognition and handling" describes a method of document image matching by performing some image processing and forming feature vectors from the pixel distributions within the document.

[0005] Disclosures of the patent and all references discussed above and in the Detailed Description of the invention are hereby incorporated herein by reference.

Summary of the Invention

[0006] The invention is a method for matching objects, with specific examples of matching documents, based on spatial layout of regions that addresses the above difficulties. It employs a shape similarity model for detecting similarity between 2D objects. The shape similarity model is general enough to encompass the individual region shape variations between members of a shape class, and yet specific enough to avoid mismatches to objects with perceptually different appearance. That is, the shape model models the change in shape of corresponding regions on objects by a set of separate affine deformations, with constraints on the transforms that are intended to capture perceptual shape similarity between objects.

[0007] Using the shape model, two objects are taken to match if one of them can be found to belong to the shape class of the other document. Specifically, the "document" matching proceeds in 4 stages, namely, (1) pre-processing, in which logical regions are automatically derived from information in the documents to be matched, (2) region correspondence, in which a correspondence is established between the regions on the documents, and (3) pose computation, in which the individual transforms relating corresponding regions are recovered, and finally (4) verification, in which the extent of spatial similarity is measured by projecting one document onto the other using the computed pose parameters.

[0008] The document matching method specifically described herein can be suitably combined with other text-based retrieval methods to enhance the capability of current document management systems. Such a document matching method has several applications. It can be used to describe document genres (such as letters, memos) based on spatial layout. Other uses of the document matching method include the clustering of documents based on similarity for purposes of document database organization.

[0009] The "object" matching method includes the following features:

1. The underlying shape model and the associated recognition method is general and is intended to capture perceptual shape similarity in a variety of 2D shapes (besides documents) that consist of regions, such as engineering drawings, MRI brain scans, video, outdoor natural scenes, house layout plans in real-estate databases, etc.
2. It is a fast method of obtaining region correspondence that avoids exponential search.
3. It has an ability to group similar shaped objects into shape categories or genres.
4. It provides a way of finding similar-looking objects under changes in object orientation, skew (rotation and shear (misfed pages)) that is fast and does not require pixel-based computations as in object image matching methods.
5. It provides an ability to retrieve documents based on spatial layout information (through query by example) which can be a suitable complement to

text-based retrieval.

6. Finally, each of the operations in object matching are computationally simple.

Description of the Drawings

[0010]

Figures 1A and 1B illustrates two similar-looking documents.
Figure 2A and 2B illustrate the similarities and differences in the layout of the logical regions of Figure 1A and 1B respectively.
Figure 3 illustrates an example illustrating region correspondence for objects.
Figure 4 illustrates a flow diagram of the document matching method.
Figure 5 illustrates an example application of the shape matching method to two diagrams.
Figures 6A illustrates a prototype document used for the comparison of other documents.
Figure 6B illustrates a document of the same category of the prototype document of 6A.
Figure 6C illustrates a document of a different category than the prototype document of 6A.
Figure 7A illustrates the projection of the document regions of the document of Figure 6B onto regions of the prototype.
Figure 7B illustrates the projection of document regions of the document of Figure 6C onto regions of the prototype.
Figure 8 illustrates a block diagram of main components for the invention.

Detailed Description of the Invention

[0011] The invention disclosed here is a method of object matching based on a shape model to capture shape layout similarity. In modeling shape similarity, objects are characterized by a collection of regions representing some logical entity of the object, such as say, a logical text region like a paragraph. The methods to obtain such regions are expected to be domain-specific and frequently involve some image pre-processing. Although the term "document" is used throughout this disclosure, it is not meant to limit the application of the invention to documents, but rather this method is intended to apply broadly to "object" matching.

[0012] The document matching method described herein generally proceeds in 4 stages afier documents containing logical regions to be matched are identified. The steps are: Pre-processing 1, in which logical regions are automatically derived from information in the documents to be matched; Region correspondence 2, in which a correspondence is established between the regions on the documents; Pose computation 3, in which the individual transforms relating corresponding regions are recovered, and finally, Verification 4 is conducted, in which the extent of spatial similarity is measured by projecting one document onto the other using the computed pose parameters. These steps will be discussed in more detail in Section B of this disclosure, below

[0013] The document matching method is based on the following shape similarity model that is intended to capture spatial layout similarity of logical regions in 2d objects and, in particular, logical regions of a document image.

A. The shape similarity model

[0014] The shape similarity model describes the characteristics of the shape class of an object M consisting of a set of m regions $R_{mi,\ i = 1\ to\ m}$. According to the shape model, the object M is said to be similar to another object I characterized by the set of regions $R_{Ij,j = 1,..n.}$ If enough pairs of corresponding regions can be found such that the shape deformations of the corresponding regions can be specified by a set of affine transforms (A_{ij}, T_{ij}) that obey the following three constraints:

1. **Direction of residual translation constraint:**

[0015] The first constraint specifies that object regions displace along a common direction, called the reference direction. That is, the direction of residual translation of corresponding regions must be the same which is denoted by:

$$\gamma_{ijy} = \gamma_{ijx} \tan\theta, \forall i,j$$

where $(\gamma_{ijx}, \gamma_{ijy})^T = \gamma_{ij} = C_{Ij} - C_{Mi}$ is the residual translation and C_{Ij} and C_{Mi} are the centroids of the regions R_{Ij} in object I and R_{Mi} of object M. When $\theta = 90$ degrees, $\gamma_{ijx} = 0$.

[0016] The direction of residual translation can be either manually specified or can be automatically derived by computing the direction of residual motion for all pairs of regions, and recording the commonality in direction using a Hough transform.

2. **Extent of translation constraint:**

[0017] The second constraint restricts the amount of displacement each of the regions can undergo to maintain perceptual similarity. The extent of residual translation of all corresponding regions is bounded by δ so that

$$|\sqrt{1+\tan^2\theta}\,\gamma_{ijx}| \le \delta$$

or equivalently:

$$|\gamma_{ijx}| \le |\delta \cos\theta|$$

$$|\gamma_{ijy}| \leq |\delta \sin \theta|$$

For $\theta = 90$, the bound on the extent of translation is given by $|\gamma_{ijy}| \leq \delta$.

[0018] Note that included in the constraint on the extent of residual translation, is the case when some regions don't move at all from their original positions while others do along a common direction. When (A_{ij}, T_{ij}) is the same for all regions j for an object i, this reduces to a rigid-body shape model.

3. Ordering of regions constraint:

[0019] The final constraint restricts the displacement of regions such that relative ordering is maintained. That is, the ordering of corresponding regions on objects with respect to the reference direction θ be the same. The ordering of regions is obtained by projecting the centroids of regions onto the reference direction using a direction of projection (orthogonal or oblique). Such a region ordering for an object can be conveniently represented by a sequence $R = (R_{j1}, R_{j2}, .. R_{jm})$. Regions of the same rank appear in this sequence ordered along the direction of projection.

[0020] The above constraints have been carefully chosen through studies that observe that such constraints perceptual shape similarity for a wide variety of objects, including faces, MRI scans, etc.

B. The method of document matching

[0021] The method of document matching disclosed in this invention involves the following stages:

1. Logical region extraction from document segments.
2. Region correspondence between the two documents to be matched using the constraints of the shape model.
3. Pose computation between corresponding regions.
4. Pose verification by projecting one of the documents onto the other using the computed pose.

1. Logical region extraction

[0022] To use the shape similarity model for document matching, a set of logical regions need to be derived. While the document matching methods admits several methods of obtaining logical regions, we chose to obtain them by a grouping algorithm that uses text segments given by a conventional text segmentation algorithm (we used a version of software in Xerox's TextBridge™ for extracting text segments). The grouping algorithm performs the following operations:

1. Text segment regions whose bounding boxes are left aligned, right aligned, or centrally aligned are noted.
2. Among the regions in Step-1, those that are vertically spaced by a small distance are retained. The distance thresholds is chosen relative to the page image size. One way to derive the threshold is to record the inter-region separation over a large training set of documents of a certain genre and record the pixel-wise separation in the image versions of documents.
3. Text regions retained in Step-2 are used to form groups of regions. The grouping is done as follows:

   a. Initially put all text segments into their own groups.
   b. For each text segment, determine the text segments that fall within the logical region distance constraint (given above). Merge all such regions into one group.
   c. Successively merge groups using step b, until the set of groups cannot be further reduced.

[0023] The above algorithm can be efficiently implemented using a data structure called the union-find data structure as described in a book by Cormen, Leisersen and Rivest entitled "Introduction to algorithms", MIT Press, 1994, to run in time linear in the number of text regions in the document.

[0024] The above algorithm has been found particularly useful for grouping consecutive paragraphs of text into single logical regions, as well as for grouping centrally aligned information such as author information in a journal article.

2. Region correspondence

[0025] The method of obtaining region correspondence is again meant for general objects, and can be easily adapted to logical regions of documents. The correspondence between logical regions is obtained by using the constraints in the shape similarity model. Thus starting with all pairs of regions on the two objects, all those pairs whose direction of residual translation is not in the specified direction θ are pruned (this is checked within a threshold to allow some robustness against segmentation errors and deviations from the shape similarity model in individual documents). Next, the extent of residual translation constraint is used to further prune the pairs. The distinct regions in the pairs on each object can now be ordered with respect to the reference direction θ. The region orderings can be denoted by the sequences R_M and R_I respectively. Using the region ordering, and collecting the set of candidate matching regions in object M for each region of object I by S_i the result can be denoted by the set sequence $S_p = (S_1, S_2, .. S_p)$ where p is the number of regions in object I that found a match to a region in object M.

The best region correspondence is taken as the longest subsequence of R_M that is also a member sequence of S_p. A member sequence of S_p is a sequence of regions with at most one region taken from the set S_i. It can be easily shown that the problem of finding the longest common subsequence (LCS) has optimal substructure property, thus admitting solutions by dynamic programming. In fact, we adapted the dynamic programming algorithm for computing an LCS described in Introduction to algorithms by T.Cormen, Leiserson, and R. Rivest, to give region correspondence by LCS using the following simple procedure:

- Let m = length of sequence R_M and n = length of set sequence S_p.
- The LCS of R_M and S_p is determined by first computing the length of LCS and backtracking from the indices contributing to the longest sequence to derive an LCS.
- The intermediate results are recorded in a dynamic programming table $c[0..m,0..n]$, where entry $c[i,j]$ denotes the length of LCS based on the prefixes of R_M and S_p of length i and j respectively. The table is initialized by $c[i,0] = 0 = c[0j]$ for all i,j.

The code is given below:

```
for i = 1 to m do
  for j = 1 to n do
    if x_i ∈ S_j
      then  c[i,j] = c[i-1,j-1]+1
      else
        if c[i-1,j] ≥ c[i,j-1]
          then  c[i,j] = c[i-1,j]
          else  c[i,j] = c[i,j-1]
```

By keeping a record of which of the three values $c[i-1,j-1],c[i-1,j],c[i,j-1]$ actually contributed to the value of $c[i,j]$ in the above procedure, we can reconstruct an LCS in linear time.

[0026] The above steps give the largest set of corresponding regions between query and prototype that move along the same direction within bounds of the shape model, have a match in individual content, and retain the spatial layout ordering specified. Although the number of possible LCS for general sequences can be exponential in the worst case, for typical spatial layouts of regions in documents, only a few distinct LCS have to be tried to discover shape similarity.

Region Correspondence Example:

[0027] Figures 3A and 3B is used to illustrate the region correspondence for a simple example. The figures are two objects to be matched some of whose regions meet the constraints of the shape similarity model. Their respective region orderings with respect to the reference direction shown in the figure, are given by *(ABCD)* and *(EFGHIJK)*. The set sequence of candidate matches is given by *({A},{B},{A},{B},{C},{},{D})*, where $S_1 = \{A\}$, $S_2 = \{B\}$ and so on. There are two LCS that have length with respective region correspondences as *{(A,E)(B,H)(C,I)(D,K)}* and *{(A,G)(B,H)(C,I)(D,K)}*. The correctness of these correspondences can be judged in the recognition stage to be described next.

3. Pose computation

[0028] Using the correspondence between logical regions, the individual region transforms can be recovered in a variety of ways including feature matching or direct region matching as mentioned in a paper entitled "Object recognition by region correspondence" in Proceedings Intl. Conference on Computer Vision (ICCV), Boston, 1995 by R. Basri and D. Jacobs. For the domain of documents, since the logical regions are rectangular, the pose parameters of interest are the four elements of the linear transform matrix A and the residual translation T. For a pair of corresponding regions R_{Mi} and R_{Ij} these are denoted by

$$A_{ij} = \begin{pmatrix} s_{1ij} & 0 \\ 0 & s_{2ij} \end{pmatrix}, \; T_{ij} = \begin{pmatrix} T_{1ij} \\ T_{2ij} \end{pmatrix}$$

where

$$S_{1ij} = \Delta x_j / \Delta x_i$$

$$S_{2ij} = \Delta y_j / \Delta y_i$$

$$T_{ij} = C_{Mj} - C_{Ij}$$

where $(\Delta x, \Delta y)$ are the width and height of the rectangular region.

4. Pose Verification

[0029] Pose verification involves determining if the two documents register under the shape similarity model. For this the computed residual translation given in the above equation is corrected such that the resulting residual translation is forced to be exactly along the reference direction and within the stated bounds on the extent of such displacement. This is done by perpendicularly projecting the point representing the computed

residual translation onto the line of direction θ and taking the point of such projection as the new residual translation T_{ijnew} for each pair of corresponding regions.

[0030] Each rectangular region R_i on object M can now be projected onto the object I to give the projected rectangular region R_i' as follows. The centroid of the region C_{Mi} is moved to the position

$$C_{Ii}' = C_{Mi} + T_{ijnew}$$

Verification is then done by seeing the extent of overlap between R_i' region and the corresponding rectangular region R_{Ij} of the correspondence pair. The verification score is given by $V(M,I)$

$$V(M,I) = \frac{\sum_{i,j} R_i' \cap R_{Ij}}{\sum_k R_{Mk} \cup \sum_l R_{Il}}$$

where $\cap$ and $\cup$ are done over the region areas.

The above formula accounts for the extent of match as measured by the extent of spatial overlap of corresponding regions, and the extent of mismatch as measured by the areas of regions that do not find the match (included in the denominator term).

C. Examples

[0031] Referring to Figure 4, a flow chart representing the four major steps of document recognition employing the invention, alter logical regions of documents to be matched are identified. Logical region extraction occurs within the first two blocks where First 1 all region pairs are formed, and then the documents are pruned 2 based on unary constraints discussed in further detail below. Region correspondence is then determined between documents 3. A pose is computed 4 for the documents and match verification is determined based a matching score. Figure 5 is another diagram depicting a more specific scenario where two documents (1 and 2) are scanned, and enter a region segmentation module 7 to establish a correspondence between the regions on the documents. A logical region grouping module 8 is then allowed form region pairs, and unary constraints are then applied to the documents in a layout shape matching module 9 resulting is a matching score between the documents.

[0032] Referring back to Figures 1A and 2A, illustrated is the logical region grouping for documents. Figure 1A shows text regions given by a conventional text segmentation algorithm. Figure 2A shows the result of logical grouping on the document image of Figure 1A.

[0033] Next, we illustrate region correspondence. Figure 6A depicts a model document by its logical regions. Figures 6B and C are two other documents only one of which is similar (Figure 6B) in spatial layout to the document in Figure 6A. This can be observed from the higher number of corresponding regions (labeled R1 through R5 in Figure 7) obtained when the document of Figure 6B is matched to Figure 6A. Here we assume a vertical reference direction, and the matching regions are indicating by identical colors. As can be seen, the poor correspondence of object in Figure 6C indicates a mismatch. This can also be seen during the verification stage where the pose parameters computed (and corrected) from region correspondences defined by overlap as the document of Figure 6A is overlayed onto the documents of Figures 6B and C. The extent of overlap in such overlay is indicated in Figures 7A and B, respectively, as R1 through R5.

D. Document matching System

[0034] The method of the invention can be carried out in a microprocessor-based system 10 as shown generally in Figure 8. The microprocessor 11 would be programmed to carry out the four main steps of the invention. A memory 12 would be utilized by the microprocessor 11 for storing document templates and tested documents during matching operations. A scanner 13 may be used to scan the test documents into the document matching system; however, as known in the art, documents may be delivered to the system via electronic networks, or the like. Results of testing can be output 14 to the user with indicating means known in the art.

[0035] The method of document matching by spatial region layout can be a useful complement to existing methods for document matching based on text keywords or pixel-wise image content. As can be seen from the examples above, the capturing of spatial layout similarity allows the matching of documents that have greater variations in pixel-wise image content. In addition, the matching method is a general formulation that can be applied to other classes of 2D objects besides documents.

Claims

1. A method of matching objects based on spatial region layout, wherein:

   objects containing logical regions to be matched are examined;
   logical regions are automatically derived from information in said obejcts;
   a correspondence is established between said regions;
   individual transforms relating said regions based on said correspondence are recovered; and
   spatial similarity between said obejcts is measured by projecting one shape onto another

obejct other using computed pose parameters.

2. The method of claim 1 wherein said method is used to match documents and is combined with a text-based retrieval method to enhance document management system capabilities.

3. The method of claim 1 wherein said method is used to describe document genres based on spatial layout.

4. The method of claim 1 wherein said method is used for clustering a plurality of obejcts based on similarity in a database.

5. The method of claim 1 wherein said method used to capture obejct similarities in a plurality of 2-dimensional obejcts that comprise of regions.

6. The method of claim 5 wherein said method groups similar shaped objects into shape categories or genres.

7. The method of claim 1 wherein said method matches shapes regardless of orientation and skew.

8. A document matching method for matching documents based on spatial region layout, comprising the steps of:

pre-processing documents in which logical regions are automatically derived from information in said documents to be matched;
determining a regional correspondence between said regions on said documents;
recovering individual transforms relating said regions based on said regional correspondence; and
verifying spatial similarities between said documents by projecting one document onto the other.

9. A system for matching documents based on spatial region layout, comprising:

pre-processing means in which logical regions are automatically derived from information in the documents to be matched;
region correspondence means in which a correspondence is established between the regions on the documents;
pose computation means in which the individual transforms relating corresponding regions are recovered; and
verification means in which the extent of spatial similarity is measured by projecting one document onto the other using the computed pose parameters.

10. A system for matching documents based on spatial region layout, comprising:

a microprocessor programmed to:

identify documents containing logical regions to be matched;
derive logical regions from information in the documents to be matched;
establish a correspondence between said regions on said documents;
recover individual transforms relating to corresponding regions; and measure spatial similarity between documents by electronically projecting one document onto the other using computed pose parameters.

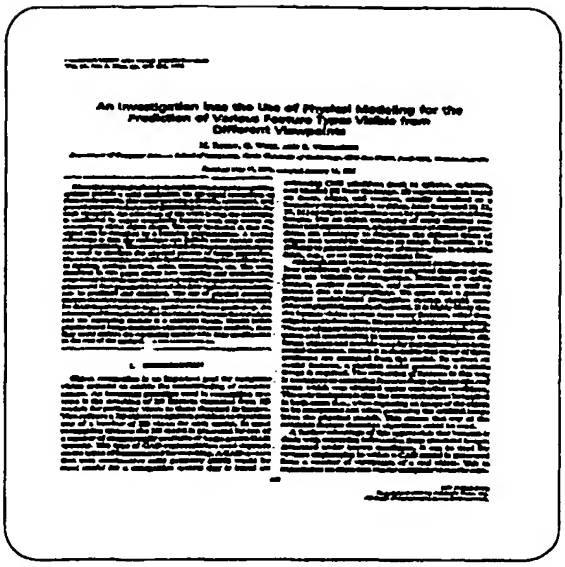

FIG. 1A

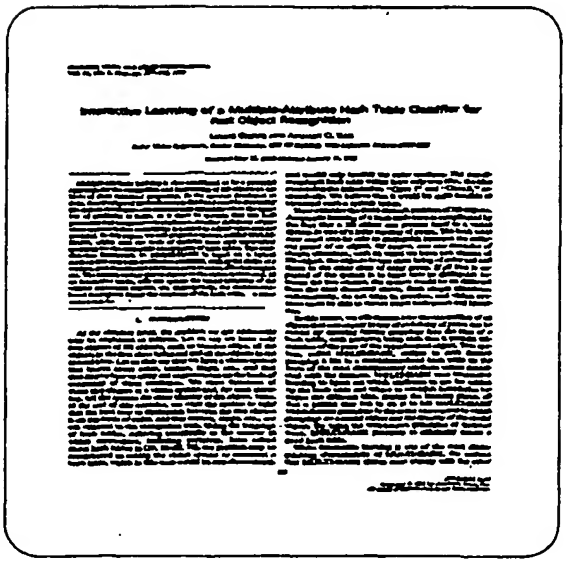

FIG. 1B

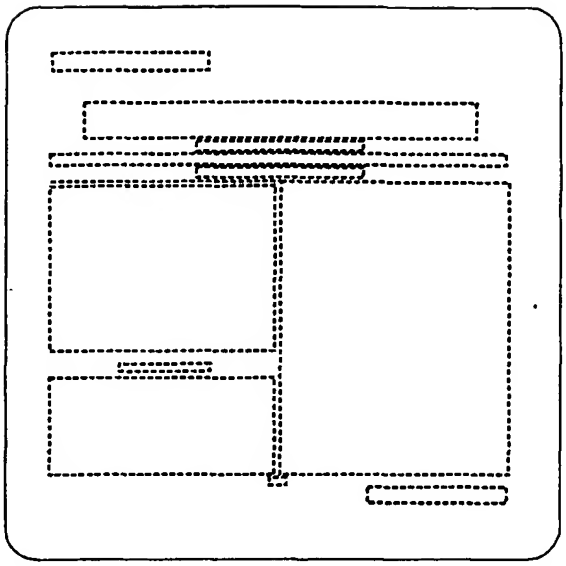

FIG. 2A

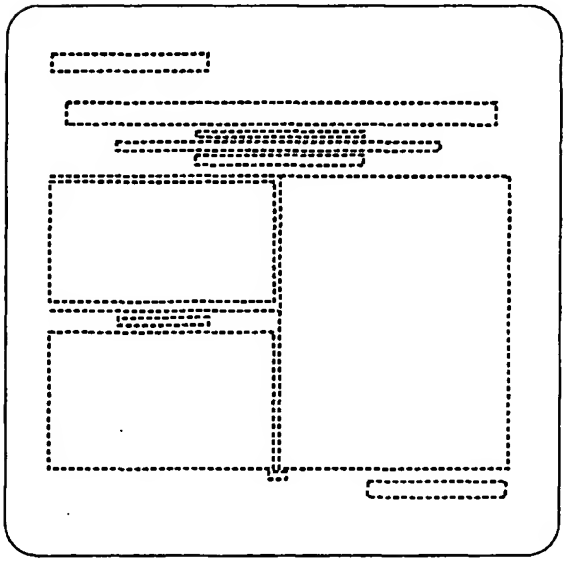

FIG. 2B

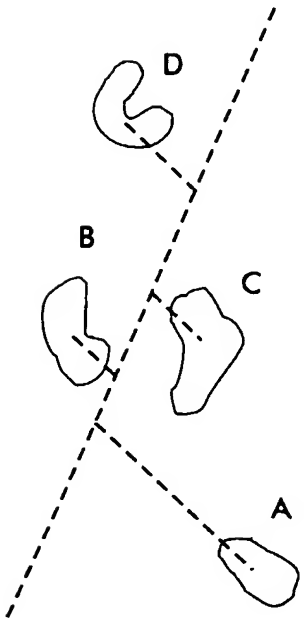

FIG. 3A

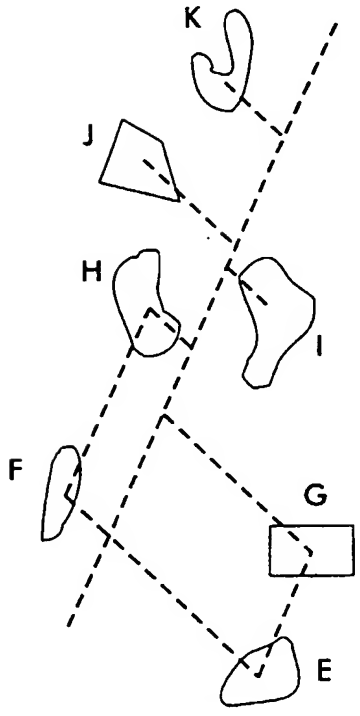

FIG. 3B

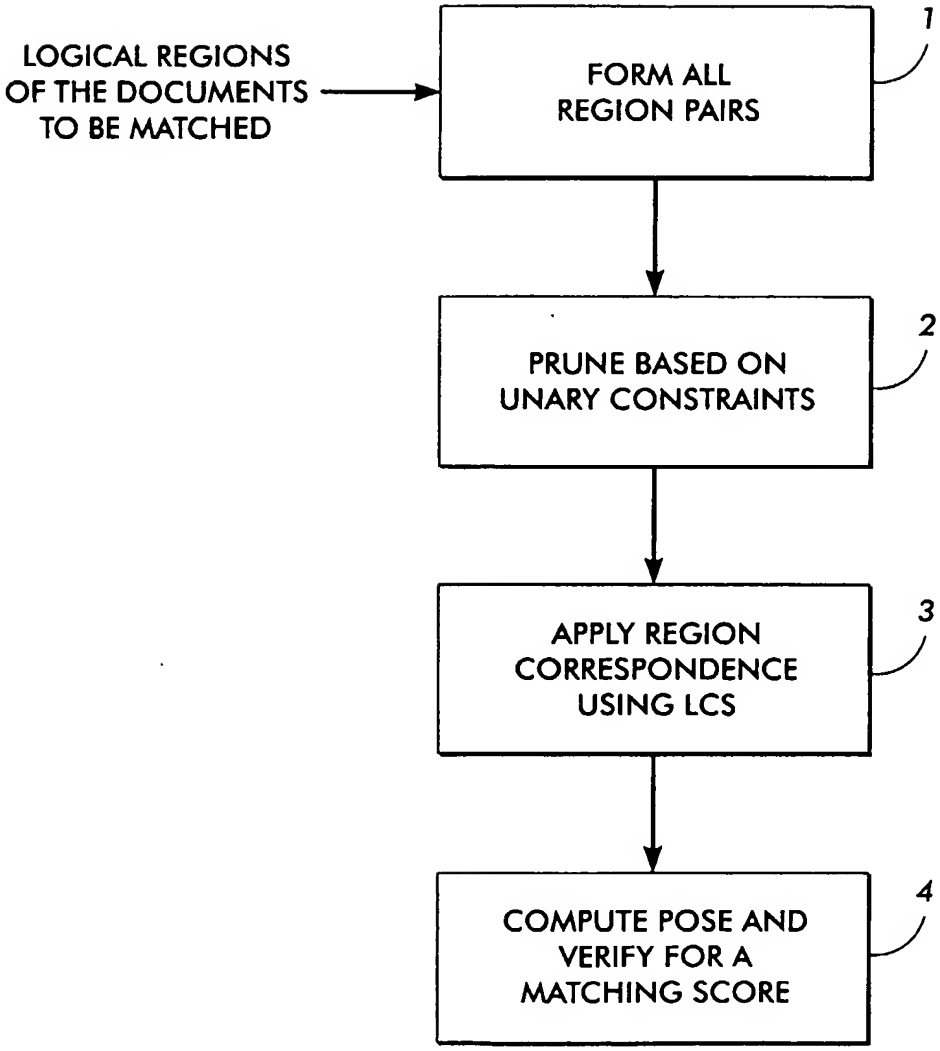
LOGICAL REGIONS OF THE DOCUMENTS TO BE MATCHED
FORM ALL REGION PAIRS
1
PRUNE BASED ON UNARY CONSTRAINTS
2
APPLY REGION CORRESPONDENCE USING LCS
3
COMPUTE POSE AND VERIFY FOR A MATCHING SCORE
4

FIG. 4

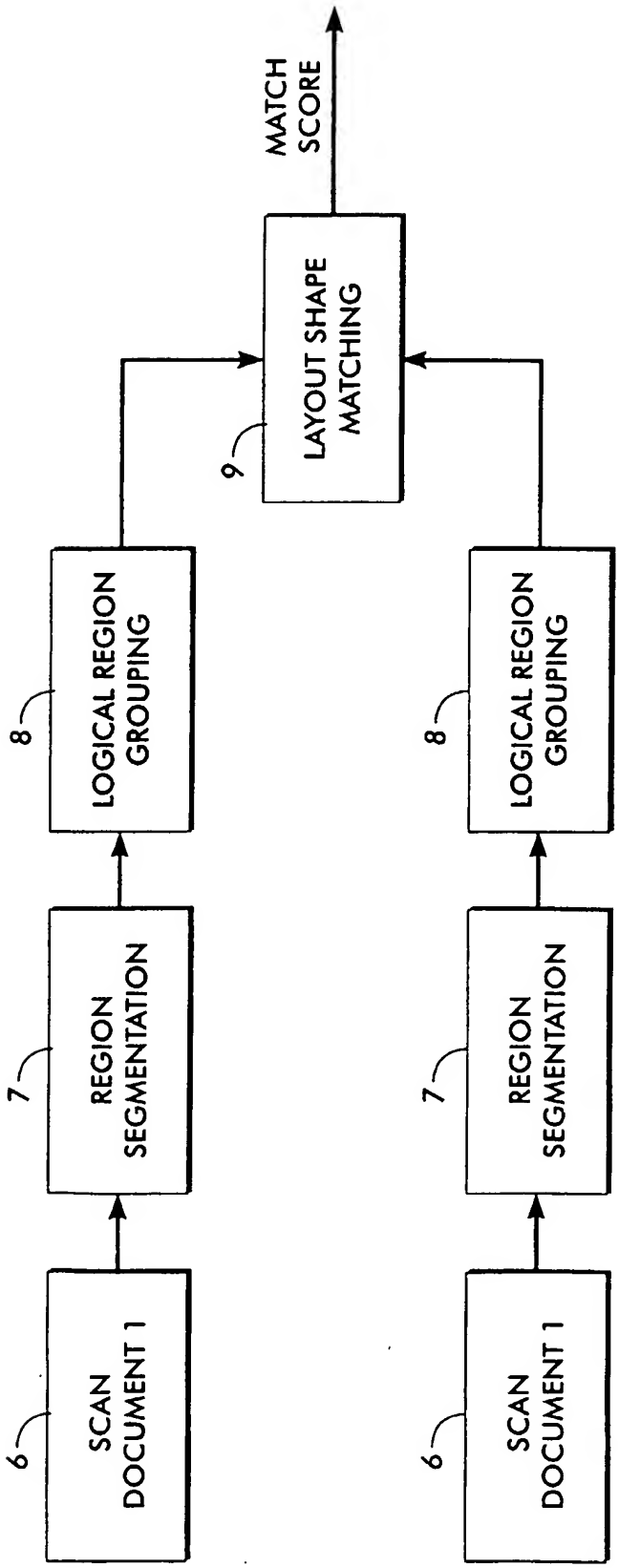
MATCH SCORE
LAYOUT SHAPE MATCHING
9
LOGICAL REGION GROUPING
8
REGION SEGMENTATION
7
SCAN DOCUMENT 1
6
LOGICAL REGION GROUPING
8
REGION SEGMENTATION
7
SCAN DOCUMENT 1
6

FIG. 5

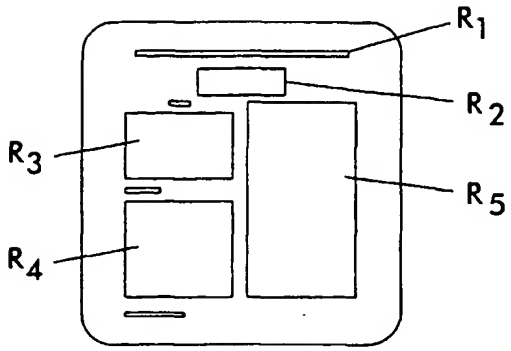

FIG. 6A

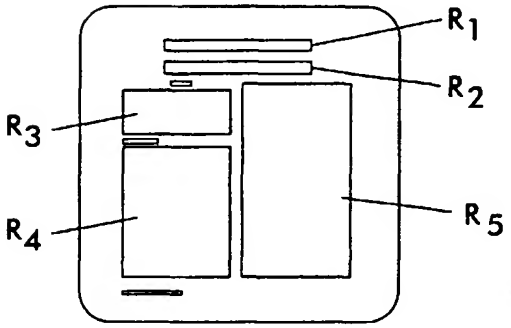

FIG. 6B

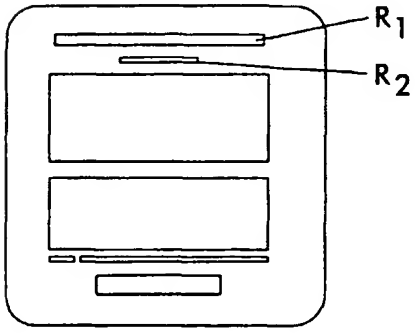

FIG. 6C

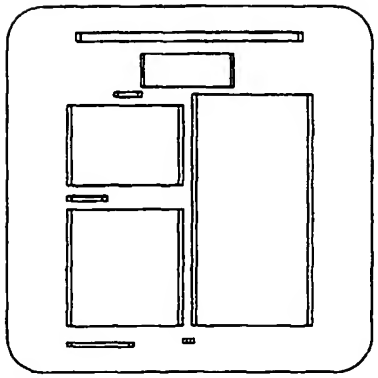

FIG. 7A

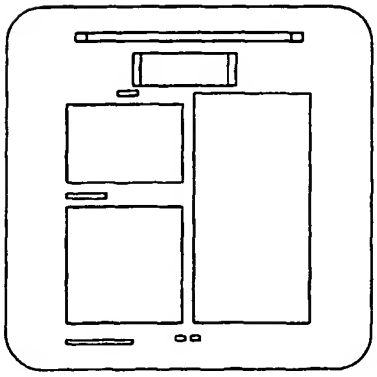

FIG. 7B

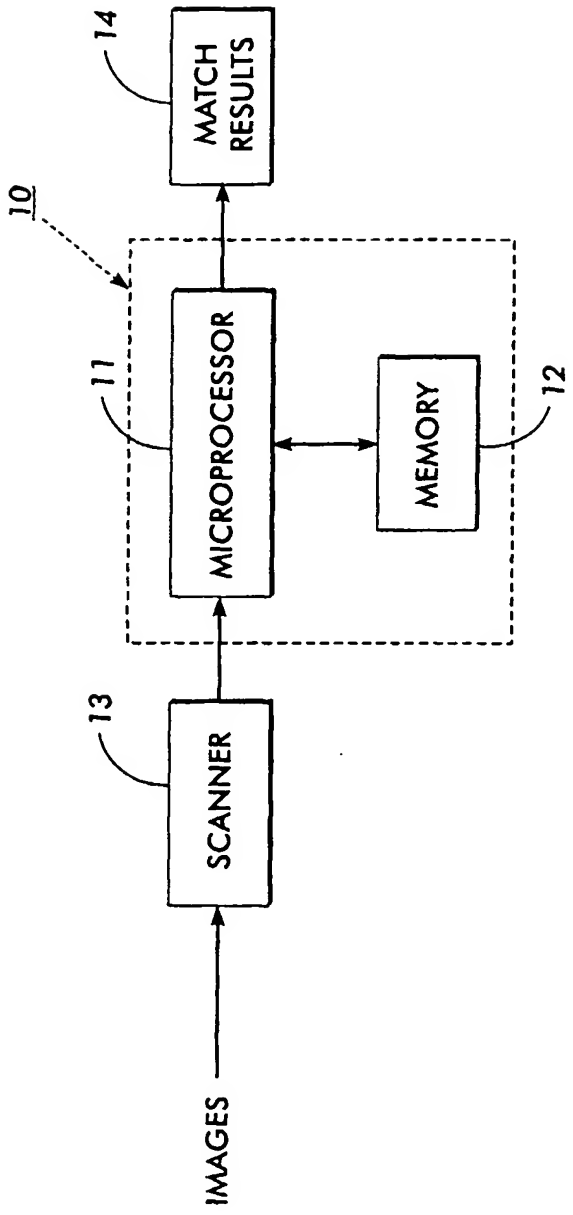
14
MATCH RESULTS
10
11
MICROPROCESSOR
MEMORY
12
13
SCANNER
IMAGES

FIG. 8

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☑ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☑ SKEWED/SLANTED IMAGES
- ☑ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.